**Project Specification**

**Finding Donors for CharityML**

Exploring the Data

| **Criteria** | **Meets Specifications** |
| --- | --- |
| Data Exploration | Student's implementation correctly calculates the following:   * Number of records * Number of individuals with income >$50,000 * Number of individuals with income <=$50,000 * Percentage of individuals with income > $50,000 |

Preparing the Data

| **Criteria** | **Meets Specifications** |
| --- | --- |
| Data Preprocessing | Student correctly implements one-hot encoding for the feature and income data. |

Evaluating Model Performance

| **Criteria** | **Meets Specifications** |
| --- | --- |
| **Question 1:** Naive Predictor Performance | Student correctly calculates the benchmark score of the naive predictor for both accuracy and F1 scores. |
| **Question 2:** Model Application | The pros and cons or application for each model is provided with reasonable justification why each model was chosen to be explored.  Please list all the references you use while listing out your pros and cons. |
| Creating a Training and Predicting Pipeline | Student successfully implements a pipeline in code that will train and predict on the supervised learning algorithm given. |
| Initial Model Evaluation | Student correctly implements three supervised learning models and produces a performance visualization. |

Improving Results

| **Criteria** | **Meets Specifications** |
| --- | --- |
| **Question 3:** Choosing the Best Model | Justification is provided for which model appears to be the best to use given computational cost, model performance, and the characteristics of the data. |
| **Question 4:** Describing the Model in Layman's Terms | Student is able to clearly and concisely describe how the optimal model works in layman's terms to someone who is not familiar with machine learning nor has a technical background. |
| Model Tuning | The final model chosen is correctly tuned using grid search with at least one parameter using at least three settings. If the model does not need any parameter tuning it is explicitly stated with reasonable justification. |
| **Question 5:** Final Model Evaluation | Student reports the accuracy and F1 score of the optimized, unoptimized, models correctly in the table provided. Student compares the final model results to previous results obtained. |

Feature Importance

| **Criteria** | **Meets Specifications** |
| --- | --- |
| **Question 6:** Feature Relevance Observation | Student ranks five features which they believe to be the most relevant for predicting an individual's’ income. Discussion is provided for why these features were chosen. |
| **Question 7:** Extracting Feature Importances | Student correctly implements a supervised learning model that makes use of the feature\_importances\_ attribute. Additionally, student discusses the differences or similarities between the features they considered relevant and the reported relevant features. |
| **Question 8:** Effects of Feature Selection | Student analyzes the final model's performance when only the top 5 features are used and compares this performance to the optimized model from **Question 5**. |